

What Is Claimed Is:

1. In a communication device coupled to an InfiniBand network and an external communication system, a method of transferring a communication
5 from the InfiniBand network to the external communication system, the method comprising:
- receiving packets for each of multiple queue pairs terminating at the communication device;
- for each of said queue pairs:
- 10 reassembling in a shared memory contents of said packets into communications to be transmitted to the external communication system; and
- maintaining an associated linked list identifying locations in said shared memory in which said communications are reassembled; and
- 15 when a communication is reassembled for a first queue pair, identifying to a transmission module a portion of a first linked list associated with said first queue pair, wherein said first linked list portion identifies shared memory locations in which said communication was reassembled.
- 20 2. The method of claim 1, further comprising:
- only after said reassembled communication is transmitted, removing said portion of said first linked list from said first linked list.
3. The method of claim 1, wherein said reassembling comprises:
- 25 as said packets are received from the InfiniBand network, queuing said contents directly into said shared memory;
- wherein said shared memory serves as receive queues for each of said

multiple queue pairs.

4. The method of claim 3, further comprising:
transmitting said reassembled communication from said shared memory;
5 wherein said shared memory serves as a transmit queue for the external
communication system.

5. The method of claim 1, wherein the external communication
system is an Ethernet network.

10

6. The method of claim 5, wherein said received packets comprise
portions of encapsulated Ethernet packets.

7. The method of claim 1, wherein said identifying to a transmission
15 module comprises transferring to the transmission module a set of pointers
identifying said reassembled communication rather than transferring to the
transmission module said reassembled communication.

8. The method of claim 1, wherein said identifying to a transmission
20 module comprises identifying to the transmission module:
a head of said portion of said first linked list; and
a tail of said portion of said first linked list.

9. The method of claim 1, wherein said maintaining an associated
25 linked list for said first queue pair comprises:
maintaining a head pointer configured to identify:
a first location in said shared memory in which contents of a first

packet are stored; and

a first entry in a shared control structure, said first entry corresponding to said first location in said shared memory; and maintaining a tail pointer configured to identify:

5 a final location in said shared memory in which contents of a final packet are stored; and

a final entry in said shared control structure;

wherein each entry in said shared control structure that is part of said first linked list, except for said final entry, identifies a subsequent entry in said shared control structure and identifies a location in said shared memory corresponding to said subsequent entry.

10. The method of claim 9, further comprising:

protecting said shared control structure by one or more of:

15 separating bits of the structure to prevent double bit errors; and

providing single error correct and double error detect protection for one or more entries in the control structure.

11. The method of claim 10, further comprising: extending the single error correct and double error detect protection to include said location in said shared memory.

12. The method of claim 1, further comprising:

managing said linked lists for said queue pairs with a shared control;

25 wherein each said location in said shared memory corresponds to an entry in said shared control; and

wherein each entry in said shared control is configured to identify:

a subsequent entry within the same linked list; and
a location in said shared memory corresponding to said subsequent entry.

- 5 13. A computer readable medium storing instructions that, when
executed by a computer, cause the computer to perform a method of transferring a
communication from an InfiniBand network to a communication system external
to the InfiniBand network, the method comprising:
for each of multiple queue pairs terminating at the communication device,
10 receiving packets;
for each of said queue pairs:
reassembling in a shared memory contents of said packets into
communications to be transmitted to the external communication system;
and
15 maintaining an associated linked list identifying locations in said
shared memory in which said communications are reassembled; and
when a communication is reassembled for a first queue pair, identifying to
a transmission module a portion of a first linked list associated with said first
queue pair, wherein said first linked list portion identifies shared memory
20 locations in which said communication was reassembled

14. The computer readable medium of claim 13, the method further
comprising:
as said packets are received from the InfiniBand network, queuing said
25 contents directly into said shared memory, wherein said shared memory serves as
receive queues for each of said multiple queue pairs; and
transmitting said reassembled communication from said shared memory,

wherein said shared memory also serves as a transmit queue for the external communication system.

15. The computer readable medium of claim 13, wherein said
5 identifying to a transmission module comprises transferring to the transmission module a set of pointers identifying said reassembled communication rather than transferring to the transmission module said reassembled communication.

16. A method of storing a communication, received from an
10 InfiniBand network, for transmission external to the InfiniBand network, the method comprising:

receiving a set of InfiniBand packets from an InfiniBand network, each said packet comprising a portion of a communication to be transmitted external to the InfiniBand network;

15 storing said communication portions in a memory shared among multiple queue pairs of the InfiniBand network, including a first queue pair through which said set of InfiniBand packets is received;

maintaining a first linked list for said first queue pair to identify locations in said memory in which said communication portions are stored; and

20 when all of said communication portions are stored in said memory, scheduling said communication for transmission from said memory.

17. The method of claim 16, wherein said storing comprises reassembling said communication portions into said communication.

25

18. The method of claim 16, wherein said scheduling comprises:
identifying to a transmission module a first entry in said first linked list

corresponding to a location in said memory in which a first portion of said communication is stored; and

identifying to a transmission module a final entry in said first linked list corresponding to a location in said memory in which a final portion of said
5 communication is stored.

19. The method of claim 16, wherein said InfiniBand packets comprise Send commands conveying said portions of said first communication.

10 20. The method of claim 16, wherein said communication is an Ethernet packet.

21. The method of claim 16, wherein said maintaining a first linked list comprises:

15 maintaining a head pointer identifying:

a first entry in a control structure; and

a first location in said memory in which a first portion of said communication is stored; and

maintaining a tail pointer identifying:

20 a final entry in said control structure; and

a final location in said memory in which a final portion of said communication is stored;

wherein said first entry is linked to said final entry by zero or more intermediate entries in said control structure, each said intermediate entry
25 corresponding to an intermediate location in said memory in which a portion of said communication is stored.

22. The method of claim 21, further comprising:
updating said first linked list to remove said first entry, said final entry and
said intermediate entries from said first linked list only after said communication
is transmitted.

5

23. The method of claim 16, wherein said maintaining a first linked list
comprises:

in a control structure, maintaining a first linked list of control entries,
wherein each of said control entries except a final control entry identifies:

10

a subsequent control entry; and

corresponding to said subsequent control entry, a location in said
memory in which data received through said first queue pair are stored.

24. The method of claim 23, wherein said control structure and said
15 memory are separate physical memory structures.

25. The method of claim 24, wherein said control structure and said
memory employ common addressing via said first linked list.

20 26. The method of claim 23, wherein said maintaining further
comprises:

maintaining a head pointer identifying a first control entry in said first
linked list and a first location in said memory; and

25 maintaining a tail pointer identifying said final control entry in said first
linked list and a final location in said memory.

27. The method of claim 23, further comprising:

identifying to a transmission module a sub-list of said first linked list of control entries, wherein said sub-list includes control entries corresponding to all locations in said memory in which portions of said communication are stored.

- 5 28. The method of claim 27, further comprising:
 removing said sub-list of control entries from said first linked list only after said communication is transmitted.

- 10 29. A computer readable medium storing instructions that, when executed by a computer, cause the computer to perform a method of storing a communication, received from an InfiniBand network, for transmission external to the InfiniBand network, the method comprising:
 receiving a set of InfiniBand packets from an InfiniBand network, each said packet comprising a portion of a communication to be transmitted external to
15 the InfiniBand network;
 storing said communication portions in a memory shared among multiple queue pairs of the InfiniBand network, including a first queue pair through which said set of InfiniBand packets is received;
 maintaining a first linked list for said first queue pair to identify locations
20 in said memory in which said communication portions are stored; and
 when all of said communication portions are stored in said memory, scheduling said communication for transmission from said memory.

- 25 30. The computer readable medium of claim 29, wherein said scheduling comprises:
 identifying to a transmission module a first entry in said first linked list corresponding to a location in said memory in which a first portion of said

communication is stored; and

identifying to a transmission module a final entry in said first linked list corresponding to a location in said memory in which a final portion of said communication is stored.

5

31. The computer readable medium of claim 29, wherein said maintaining a first linked list comprises:

in a control structure, maintaining a first linked list of control entries, wherein each of said control entries except a final control entry identifies:

10 a subsequent control entry; and

corresponding to said subsequent control entry, a location in said memory in which data received through said first queue pair are stored; maintaining a head pointer identifying a first control entry in said first linked list and a first location in said memory; and

15 maintaining a tail pointer identifying said final control entry in said first linked list and a final location in said memory.

32. The computer readable medium of claim 31, wherein the method further comprises:

20 identifying to a transmission module a sub-list of said first linked list of control entries, wherein said sub-list includes control entries corresponding to all locations in said memory in which portions of said communication are stored.

33. The computer readable medium of claim 32, wherein the method further comprises:

25 removing said sub-list of control entries from said first linked list only after said communication is transmitted.

34. A computer readable medium containing a data structure configured for simultaneously queuing contents of packets as they are received from an InfiniBand network and storing communications reassembled from said contents for transmission external to the InfiniBand network, the data structure comprising:

5

a shared memory, comprising:

for each of a plurality of InfiniBand communication connections, memory buckets for storing contents of packets received on the communication connection;

10

wherein said contents of said packets are reassembled into communications in said memory buckets as said contents are stored;

a shared control, comprising:

for each of said communication connections, a linked list of control entries, wherein each said control entry is configured to identify:

15

a subsequent control entry in said linked list; and

in said shared memory, a memory bucket corresponding to said subsequent control entry;

wherein said shared memory and said shared control are configured for access by:

20

an InfiniBand receive module configured to receive said packets;

and

a transmit module configured to transmit said communications external to the InfiniBand network.

25

35. The computer readable medium of claim 34, wherein said shared control further comprises:

a linked list of free entries, wherein each said free entry is configured to identify, in said shared memory, a free memory bucket.

36. The computer readable medium of claim 34, wherein said shared
5 control is protected by one or more of:
separating bits of the structure to prevent double bit errors; and
providing single error correct and double error detect protection for one or
more control entries in the shared control.

10 37. The computer readable medium of claim 36, wherein a control
entry in said shared control is further protected by extending the single error
correct and double error detect protection to include the identity of a memory
bucket in said shared memory.

15 38. An apparatus for storing data received from an InfiniBand network,
for transfer to an external communication system, comprising:
an InfiniBand receive module configured to receive packets from a
plurality of InfiniBand communication connections;
a transmit module configured to transmit communications to a
20 communication system external to the InfiniBand network via one or more
outbound ports;
a memory shared between said InfiniBand receive module and said
transmit module; and
a control, shared between said InfiniBand receive module and said
25 transmit module, for maintaining a linked list for each of said communication
connections and for each of the outbound ports.

39. The apparatus of claim 38, further comprising, for each said communication connection:

a head pointer identifying a head of said linked list for said communication connection; and

5 a tail pointer identifying a tail of said linked list for said communication connection.

40. The apparatus of claim 38, further comprising, for each of the outbound ports:

10 a head pointer identifying a head of said linked list for the outbound port; and

a tail pointer identifying a tail of said linked list for the outbound port.

41. The apparatus of claim 40, further comprising, for each outbound queue of each of the outbound ports:

15 a head pointer identifying a head of said linked list for the outbound queue; and

a tail pointer identifying a tail of said linked list for the outbound queue.

20 42. The apparatus of claim 38, wherein said communication connections are queue pairs.

43. The apparatus of claim 38, wherein said communication connections are virtual lanes.

25

44. The apparatus of claim 38, wherein the external communication system comprises an Ethernet network.

45. The apparatus of claim 38, wherein the external communication system comprises a SONET network.